**Peer Review File**

Article information: http://dx.doi.org/10.21037/atm-20-413

**Reviewer:**

In the manuscript, the authors have shown that they find a prognostic nomogram that may predict the 3- and 5-year OS for patients with TNBC with histology of IDC. The findings are well-structured and have potential clinical application. Some minor points need to be clarified/modified to improve its scientific quality.

Comment 1: Table 2 only lists the significant clinicopathological data in the univariate and multivariate analyses, but why other pathological data such as family history and surgery type are not listed? What is the clinical outcome?

Reply 1: Answer: Agree with the comments and thank you for your reminder. We previously ignored to list the factors without significant. We added all the other clinical and pathological data in Supplementary table 1 and Table 2 in our manuscript in this revised version. Please see page 20-21, line 330-334. Revised text is highlighted in red color. Please see page 9, line 148-154.

Changes in the text: In the univariate analysis, the prognostic factors with p-value < 0.1 in the cohort were as follows: age, T stage, N stage, histological grade, E-cadherin status for both DFS and OS; and menopausal status, for DFS. See Supplementary table 1 for more information. The multivariate analysis found that N stage ($p < 0.001$ for both DFS and OS) and E-cadherin expression status ($p < 0.001$ for DFS; $p = 0.029$ for OS) were independent risk factors for both DFS and OS. Age ($p = 0.009$ for OS) and T stage ($p = 0.039$ for OS) were independent risk factors for OS (Table 2). Since age and T stage have clinical significance, we incorporated all these four factors into the following nomograms for DFS and OS.

Pleases refer to Supplementary table 1 (page 20, line 330-331) and Table 2 (page 21, line 333-334) in the revised manuscript.

Comment 2. The article mentioned the data of the TCGA database were used to verify the results. Why wasn't a nomogram? Moreover, only the accuracy of the model's 3-year OS has been verified. Why is the accuracy of the 5-year OS not verified?

Reply 2: For the validation of TCGA database, we created a nomogram for predicting OS using these data. We added this result in this revised manuscript as a supplementary figure 1. Please see

page 23, line 358-362. Also, we added calibration curve external validation of 5-year OS using TCGA data (supplementary figure 2a and 2b). Please see page 24, line 364-367. We described the result in the text (page 10, line 165-168). However, as the sample size is small, we discussed the limitation of the validation of long-term OS of this cohort of patients (page 12, line 212-215).

Changes in the text: ***Validation of predictive accuracy of the nomogram for OS***: The nomogram created from this cohort of patients is shown in Supplementary figure 1. The C-index of the nomogram for predicting OS was 0.843. The calibration curves for the probability of 3-year and 5-year OS were shown in Supplementary figures 2a and 2b. As data were not enough in terms of analyzing DFS in the TCGA database, we did not perform this validation.

***Discussion:*** Although the sample size of the validation cohort of TCGA datasets was small, the results of the calibration curves suggested that this nomogram was useful for predicting the short-term overall survival of TNBC patients. A larger sample size of the validation sets was required for the long-term survival results.

Pleases refer to Supplementary figure 1 (page 23, line 358-362) and 2 (page 24, line 364-367) in the revised manuscript.

Comment 3. DFS is not evaluated in this article. It is recommended that the authors evaluate DFS in TNBC patients and combine the nomogram, OS and DFS to analyze its clinical application value.

Reply 3: Agree. We evaluated the DFS in our cohort of patients, and also did internal validation. We combined the nomograms to analyze its clinical practice. Please see page 8-9, line 141-161.

Changes in the text:

***Kaplan-Meier evaluation of DFS and OS***

The median follow-up time was 70.73 months (range, 7.20-95.93 months). The 3- and 5-year survival rates were 81.0% and 76.5% for DFS, and 86.5% and 81.1% for OS, respectively. Of the 242 study patients, 50 (20.7%) died during the follow-up period. Kaplan-Meier plots of DFS and OS for TNBC patients grouped according to age, T stage, histological grade, N stage, and E-cadherin status are shown in Figures 1 and 2.

***Independent prognostic factors***

In the univariate analysis, the prognostic factors with p-value < 0.1 in the cohort were as follows: age, T stage, N stage, histological grade, E-cadherin status for both DFS and OS; and menopausal

status, for DFS. See Supplementary table 1 for more information. The multivariate analysis found that N stage ($p < 0.001$ for both DFS and OS) and E-cadherin expression status ($p < 0.001$ for DFS; $p = 0.029$ for OS) were independent risk factors for both DFS and OS. Age ($p = 0.009$ for OS) and T stage ($p = 0.039$ for OS) were independent risk factors for OS (Table 2). Since age and T stage have clinical significance, we incorporated all these four factors into the following nomograms for DFS and OS.

***Prognostic nomogram for DFS and OS***

Figure 3 illustrates the prognostic nomogram for the 3-year and 5-year DFS (a) and OS (b), as generated by the factors in the primary cohort. The calibration plots for the probability of DFS and OS showed a relatively high level of consistency between the actual observed outcome and the outcome that was predicted by the nomogram in the internal validation. The predicted accuracy for DFS and OS, as measured by the C-index, was 0.798 (Figures 4a and 4b) and 0.821 (Figures 4c and 4d).

Pleases refer to Figure 1 (page 21, line 336-339), 2 (page 21-22, line 341-344), and 3 (page 22-23, line 346-351), 4 (page 23, line 353-356).

Comment 4. The authors should discuss the clinical applicability of the nomogram, and any strategies for further optimization.

Reply 4: we add the applicability of the nomogram and how to perform further optimization in the discussion. Please see page 12-13, line 216-234.

Changes in the text: As TNBC stands out for its aggressive behavior, new biomarkers (26) and treatment entities for controlling TNBC are being explored, and many promising results have been found. In a prospective, multicenter study which investigated the addition of pembrolizumab to neoadjuvant chemotherapy for early TNBC patients, a significantly higher pathological complete response (64.8%) was found in the pembrolizumab-chemotherapy group as compared with the placebo-chemotherapy group (51.2%) (27). Another study designed and validated novel nitrogen-based chalcone analogs, which could induce a reversal of EMT by upregulating the E-cadherin (28). Using this model, we may use hypothesize that it is desperately important for patients with lower DFS or OS to try to incorporate new modalities other than traditional therapy to manage the disease. For instance, a 45-year-old patient with TNBC with IDC histology, who had a tumor of 4 cm in size, one positive axillary lymph nodes, and negative E-cadherin expression may have a probability of 82% and 72% for OS and 55% and 50% for DFS at 3 and 5 years after diagnosis, respectively. That is to say, even though the patient is in a relatively early stage, there would be

half the risk for them to suffer tumor progression five years after the surgery. More aggressive or newer agents for her was needed.

This study is limited by its small sample size and retrospective, single-center nature; therefore, further external validation is required. Additionally, data of other pathological prognostic parameters such as epidermal growth factor receptor (EGFR), androgen receptor (AR), tumor-infiltrating lymphocytes (TILs), and the breast cancer susceptibility genes (BRCA) mutation that are potentially related to prognosis are lacking. The establishment and validation of nomograms that incorporate newly and specific markers in a large cohort of TNBC patients need to be pursued.

Comment 5. Fixing inconsistent formatting of the letter “P”, italic or not should be consistent. Punctuation and space are not used correctly. For example, there should be a space before and after "<" and "=" on the results part.

Reply 5: we fixed the made the formatting of letter “*p*” consistent and all the wrong punctuation and space were corrected through out the manuscript.

Comment 6. Please kindly check the abbreviations carefully. Some abbreviations do not give the full name for the first time, such as IDC which appears in Background.

Reply 6: we checked the abbreviations and listed all of the full name for the first time. Please see page 5, line 74, page 7, line 114 and, page 6, line 98-99.

Changes in the text:

Infiltrating duct carcinoma (IDC);

cadherin 1, type 1, E-cadherin (epithelial) (CDH1);

American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP).

Comment 7. Polish your language. Some grammatical errors are found in the present manuscript.

Reply 7: we asked AME Editing Service (http://editing.amegroups.com/) for assistance and performed a thorough proofreading to eliminate the typo and grammatical errors in this revised manuscript. Attachment here is the editorial certificate.

# EDITORIAL CERTIFICATE

## AME Editing Service

This document certifies that
the manuscript listed below was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at AME Publishing Company.

**Manuscript ID:**
AES-CC9-200322

**Manuscript Title:**
Development and validation of a prognostic nomogram for patients with triple-negative breast cancer with histology of infiltrating duct carcinoma

**Author:**
Mingwei Ma

**Date Issued:**
March 26, 2020

**Price:** PAID
2240CNY / 1501-3500 words

AME Publishing Company